DOCUMENT RESUME

ED 430 390 FL 025 803

AUTHOR Johnson, Marysia
TITLE Native Speakers' Perception of the Nature of the OPI Communicative Speech Event.
PUB DATE 1999-00-00
NOTE 30p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Attitudes; *Interviews; *Language Tests; *Native Speakers; *Oral Language; *Second Languages; Speech Skills; Standardized Tests; Test Construction; Test Validity; *Verbal Tests
IDENTIFIERS *Oral Proficiency Testing

ABSTRACT

A study investigated the Educational Testing Service's claim about the conversational nature of the Oral Proficiency Interview (OPI) from the perspective of native speakers of the target second language. Eight subjects listened to 16 randomly-selected OPI communicative speech events, and their perceptions were measured using a semantic differential instrument. Analysis of the data indicate that with few exceptions, native speakers did not differ in their judgments of the nature of the OPI communicative speech event; they found that the OPI does not test speaking ability in the real-life context of a conversation as it claims to. The OPI tests speaking ability in the context of two interview types: a very formal interview that has many features of a survey research interview, based on the behaviorist theory of stimulus and response, and a more conversational type of interview that has many features of a sociologic interview. It is concluded that the findings raise questions about the validity of the OPI testing instrument. (Author/MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 430 390

Native speakers' perception of the nature of the OPI communicative speech event

Marysia Johnson
Assistant Professor
Department of English
Arizona State University

This study was conducted to investigate the Educational Testing Service's claim about conversational nature of the Oral Proficiency Interview (OPI) from an "outside" perspective. That is, this study investigated native speakers' perception of the OPI communicative speech event. Eight participants listened to 16 randomly selected OPIs, and their perceptions of the OPI were measured using the semantic differential instrument. A two-way analysis of variance was used to analyze the data. The results of the study show that, with a few exceptions, native speakers do not differ in their judgments of the nature of the OPI communicative speech event. In the native speaker's judgment, the OPI does not test speaking ability in the real-life context of a conversation as it claims to measure (ETS, 1989). The OPI tests speaking ability in the context of two types of interviews: a very formal type of interview that exhibits many features of a survey research interview, which is based on the behaviorist theory of stimulus and responses, and a more conversational type of interview that exhibits many features of a sociolinguistic interview (Mishler, 1986). These findings raise a question about the validity of the OPI testing instrument.

FL 025 803

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Marysia Johnson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Native speakers' perception of the nature of the OPI communicative speech event

This study was conducted to investigate the Educational Testing Service's claim about conversational nature of the Oral Proficiency Interview (OPI) from an "outside" perspective. That is, this study investigated native speakers' perception of the OPI communicative speech event. Eight participants listened to 16 randomly selected OPIs, and their perceptions of the OPI were measured using the semantic differential instrument. A two-way analysis of variance was used to analyze the data. Where the ANOVA F ratio was significant, a Student Newman Keuls post hoc statistical procedure was performed. The results of the study show that, with a few exceptions, native speakers do not differ in their judgments of the nature of the OPI communicative speech event. In native speakers' judgment, the OPI does not test speaking ability in the real-life context of a conversation as it claims to measure (ETS, 1989). The OPI tests speaking ability in the context of two types of interviews: a very formal type of interview that exhibits many features of a survey research interview, which is based on the behaviorist theory of stimulus and responses, and a more conversational type of interview that exhibits many features of a sociolinguistic interview (Mishler, 1986). These findings raise a question about the validity of the OPI testing instrument.

INTRODUCTION

Advocates of the Oral Proficiency Interview (OPI) claim that "a well-structured oral proficiency interview tests speaking ability in a real-life context—a conversation. It is almost by definition a valid measure of speaking ability" (ETS, 1989). However, much of how everyday conversation works is deceptively obvious so that people studying and testing language often overlook fundamental characteristics of conversation and in the process violate them. It is precisely on these grounds that van Lier (1989) has challenged the ETS's claim that it measures speaking ability in the context of a conversation. Van Lier was the first to pose the question: "Is it really a conversation?" (Van Lier, 1989, p. 494).

Since van Lier's original question, many researchers have tried to investigate the nature of the OPI speech event utilizing various discourse analysis methods (Johnson, 1997; Johnson and Tyler, 1998; Young, 1995; Lazarton, 1992). Not much research, however, has been done in the area of native speakers' perception of the OPI communicative speech event. This paper reports on the findings of the research whose purpose was to provide some answers to the following question:

Do native speakers (i.e., testers and non-testers) differ in their judgement of the nature of the OPI communicative speech event?

THE ORAL PROFICIENCY INTERVIEW

OVERVIEW

The Oral Proficiency Interview (OPI) is a widely used instrument for assessing second and foreign language speaking ability within the U.S. government institutions such as the Foreign Service Institute (FSI), and the Defense Language Institute (DLI), and nongovernment institutions such as the Educational Testing Service (ETS), and the American Council on the Teaching of Foreign Language (ACTFL).

In the OPI, which is based on the ACTFL/ETS/ILR scale and speaking level descriptions, the examinee converses face to face with one or two trained testers on a variety of topics for 10 to 30 minutes. The elicited sample is then rated on a scale ranging from 0 (no functional ability) to 5 (proficiency equivalent to that of a well-educated native speaker).

It is estimated that several thousand OPI tests are administered each year. Frequently, entire professional careers, future job assignments, pay-increases, and entrance or exit from college language programs depend on the rating obtained during the oral interview.

STRUCTURE OF THE OPI

The OPI has both a general and a level-specific structure as described in the *ILR Handbook on Oral Interview Testing* (Lowe 1988), and in Clark and Clifford (1988). The OPI consists of four phases: *Warm-up*, *Level Check*, *Probes*, and *Wind-Down*.

The *Warm-up* phase consists of social courtesies at a level that is very easy for the candidate. There are three purposes for the *Warm-up*: (1) putting the candidate at ease; (2) reacquainting the candidate with the language (if necessary); and (3) giving testers a preliminary indication of the candidate's level. This preliminary indication must be confirmed in the next phase of the interview, the *Level Check*.

The purpose of the *Level Check* is to find out the candidate's highest sustainable level of speaking proficiency. In the *Level Check* phase, testers have the candidate perform the tasks assigned to a given level. When the candidate successfully passes the *Level Check*, his/her performance provides a floor for the rating. The next phase- *Probes* - aims at finding the ceiling.

The purpose of the *Probes* phase is to show the tester(s) whether the candidate has reached his/her

highest level of speaking proficiency. To probe, testers have the candidate attempt to perform a task or tasks one level above the level of *Level Check*. The *Level Check* and *Probes* are interwoven, so that the candidate is being alternately stressed and relaxed, and not constantly pushed ever higher.

The last phase of the general structure of the interview, the *Wind-Down*, is intended to leave the candidate with a feeling of accomplishment. It also gives testers a last chance to check any aspect of the candidate's speaking ability that may have been incompletely assessed.

ELICITATION TECHNIQUES

To obtain a ratable sample (i.e., the sample to which a rating can be assigned), testers must make sure that not only general, but also level specific requirements have been fulfilled. Level specific requirements include a series of tasks and functions that are assigned to a given level. To elicit level specific tasks and functions, testers use questions and role-play situations.

A variety of question types constitute the core of the OPI elicitation procedures. A given set of question types is recommended for a particular level or levels of speaking proficiency. Thus, for instance, for *Level 0+*: ***Yes/No*** and ***Choice Questions*** are recommended; for *Level 1* and *2*: Information ***Questions***; for *Level 3*, *4* and *5*: ***Hypothetical*** and ***Supported Opinion*** questions are required. The examples of these question types are presented below (Lowe 1988):

***Yes/No Questions*:**

Do you live in Washington?

Choice Questions:

Would you like tea or coffee?

Information Questions:

What did you do last summer?

Hypothetical Questions:

If you were the Prime Minister, what would you

do to improve the economic situation in your country?

Supported Opinion Questions:

Why are you against this type of policy?

LITERATURE REVIEW

CRITICAL ANALYSIS OF THE OPI

A group of researchers (Bachman, 1988, 1990; Savignon, 1985; Bachman & Savignon, 1986; Lantolf & Frawley, 1985, 1988) working in the field of language testing and teaching have voiced strong criticism of the OPI that centers primarily on the issue of the OPI's validity, and the theory of proficiency the OPI claims to represent.

Another critic of the OPI, van Lier (1989) also calls for a thorough investigation of the OPI's construct validity. However, contrary to the authors mentioned above, he does not insist on developing an external criterion (i.e., a theoretical framework) against which the construct of the OPI should be evaluated. He calls instead for a thorough examination of the OPI from *within*. He advocates an ethnographic approach to determine what kind of speech event the OPI is, to find the answer to his original question: "Is it really a conversation?"

Van Lier finds it difficult to accept that the OPI represents instances of natural conversation because the ultimate goal of the OPI is to elicit a ratable sample, and not to conduct a conversation. Agreeing with the Jones and Gerard (1967) model of dyadic interaction, van Lier points to different distributions of rights and duties in interview and in conversation. An interview is characterized by asymmetrical contingency (i.e., the interviewer has a predefined plan and conducts the interview to execute the plan). In contrast, friendly, everyday conversation is based on mutual contingency with equal distributions of rights and duties. Van Lier, thus, finds it unclear how the OPI might accommodate these mutually exclusive types of contingency. Van Lier hypothesizes that the OPI represents a pseudo-social event, which will vary in important ways from natural conversation. If the OPI does not measure speaking ability in the form of conversation, as it claims to measure, then the users of the system may be misled about candidates' ability to actually carry on real life oral interactions.

In order to empirically investigate van Lier's hypothesis, Johnson (1997) analyzed the audio recordings of 35 official OPIs for features associated with three previously investigated speech events—spontaneous conversations, highly controlled interviews, and teacher-fronted discourse. This analysis revealed that the OPI appears to be a unique speech event, which has its own unique norms and rules. Johnson & Tyler (1998) and Tyler & Johnson (in press) expanded upon Johnson's (1997) research study by

investigating some aspects of conversational involvement and interlocutor responsiveness contained in the OPI model. The current study reports on Johnson's (1997) research study whose aim was to obtain some information about native speakers' perception of the OPI speech event (i.e., to investigate the ETS's claim about the conversational nature of the OPI speech event from an "outside" perspective).

CONVERSATION AS A SPEECH EVENT

As noted above, supporters of OPI have argued that it is a valid measure of speaking ability because it represents a real-life context—a face-to-face conversation. However, this assertion can only be maintained if careful analyses of Oral Proficiency Interviews show that they contain the established features of natural conversation. I turn now to a consideration of the research, which has examined these features.

The various works of Schegloff & Sacks (1973), Sacks, Schegloff & Jefferson (1974, 1978) and Schegloff, Jefferson & Sacks (1977) have firmly established the highly organized and locally managed system operating within conversation. Key features of the locally managed system are systematic turn-taking mechanisms and adjacency pairs. The local management system accounts for patterns of stable and recurrent actions responsible for creating order in conversation.

Examination of everyday conversation reveals that it is produced on a turn-by-turn basis. Although the turn-taking is normally accomplished smoothly, virtually no aspect of it is specified in advance. Turn size, turn order, and turn distribution are not fixed. In other words, at the beginning of a conversation the participants do not know and cannot accurately predict how much any one participant will contribute, in which order participants will talk, or how frequently any one participant will talk. Neither is the content of a participant's remarks specified in advance. Acc rding to Sacks et al. (1974), the turn-taking organization for natural conversation, by contrast with other speech exchange systems such as interviews, "makes no provision for the content of any turn, nor does it constrain what is (to be) done in any turn" (p. 710). Indeed, the unplanned nature and unpredictable outcomes constitute primary characteristics of natural conversation.

Another salient feature of everyday conversation are spontaneously created and negotiated topics. Although researchers have failed to agree upon an operational definition of topic, topic can be regarded as a "pre-theoretical notion of what is being talked about" (Brown & Yule, 1983, p. 71) "through some series of turns at talk" (Schegloff, 1979, p. 27). In natural conversation, topic is negotiated, and topical coherence is "constructed across turns by collaboration of participants" (Levinson, 1983, p. 313). We expect topic to

emerge spontaneously, and as Ochs points out "is relatively unplanned and locally managed" (Ochs, 1970, p. 58).

INTEVIEW AS A SPEECH EVENT

An interview is considered to be a prominent research method in the social and behavioral sciences. Schiffrin (1994) and Mishler (1986) distinguish between different types of interviews such as survey research interviews and sociolinguistic interviews.

A survey research interview undergoes major scientific scrutiny in Mishler's (1986) *Research Interviewing: Context and Narrative*, which leads to the establishment of several basic characteristics of that type of interviewing. In Mishler's opinion, a survey research interview is strongly embedded in the behaviorist theory of stimulus and response where an interview is viewed as a *verbal exchange* rather than a *form of discourse*.

In a survey research interview, questions and answers are regarded as stimuli and responses. All "extraneous material" is suppressed in order that the findings may be generalized to a larger population. This attempt to emulate positivistic, scientific research leads to interpreting each question and response in isolation (i.e. independent of particular features of context). The context is not viewed as an important factor influencing participants' interaction. The role of the interviewer is to become an expert in stimulus sending, so that the interviewee may become the ideal response-emitter.

Moreover, Mishler (1986) points to another typical feature of survey research interviews— the asymmetrical distribution of power. This asymmetry is evident in the interviewer's exclusive control over who will speak, when, and for how long (i.e., turn-taking), what topics are discussed, and what is relevant and not relevant to the interview.

In contrast to survey research interviews, sociolinguistic interviews allow for a variety of different genres, such as narratives or descriptions, outside a question-answer format (Schiffrin, 1994). Interviewers are trained to avoid the question-answer format and to elicit different types of talk similar to casual conversation. The role of the interviewer and interviewee in sociolinguistic interviews is also less rigidly defined. Although the asymmetrical distribution of power still exists in sociolinguistic interviews, the interviewee is allowed to change roles (i.e., ask questions of the interviewers), change and initiate topics, and have greater control over who holds the floor. Different speech events that are allowed to emerge

contribute to a better sense of cooperation and solidarity between participants (i.e., asymmetry of power is less evident in sociolinguistic interviews).

METHODS

SUBJECTS

For this research study eight participants were selected. Four of the selected participants are OPI English testers at the Defense Language Institute, and the remaining four are "naive" native speakers of English. The selection of the 8 participants was determined on the basis of several criteria, including age, education, and the willingness to participate in the study. The participants were divided into two group: Testers and Non-Testers. Within each group there were two females and two males. The age of the participants ranged from 35 to 55. All participants have a Masters Degree either in the field of TESOL, education, or computer science. All have some background in studying foreign languages at the gradual level.

MATERIALS

For the purpose of the study, 16 OPIs (i.e., four OPIs per base level) were randomly selected from 35 transcribed and coded OPIs. These 35 OPIs were also randomly selected from the pool of one of the U.S. government agencies. The OPIs were conducted in English by the OPI testers. All 35 OPIs were audiotaped. Following the testing policy of this agency, the OPIs were conducted over the telephone from the testing headquarters in Washington DC. A similar policy of substituting a telephonic OPI for a face-to-face OPI is a common practice in all U.S. government language institutions.

For each randomly selected OPI, the author prepared 3 SD forms—one for the *Warm-up*, one for the *Level Check*[1], and one for the *Wind Down*. The reason for conducting an SD for each individual phase of the OPI was to determine whether the specific phases of the OPI are perceived differently. That is, one speech event may be associated with the *Level Check* phase of the OPI and a different one with the *Wind down* phase; for instance the *Wind down* may be more "conversation-like" and *Level Check* more "interview-like." Also, it may be the case that a *Level 4 Level Check* may exhibit different speech event characteristics than a *Level 2 Level Check* phase of the OPI.

---

[1] In this study, the *Level Check* term refers to **both** the *Level Check* and the *Probe* phases. That is, for the purpose of this study, the *Level Check* includes both the *Level Check* and the *Probes* phases.

In this study, testers' and non-testers' perceptions/judgments of the OPI speech event are measured using the *semantic differential* instrument (Osgood, 1957). The semantic differential instrument has been used in many language attitude and motivation research studies (Gardner & Lambert, 1972). The well established validity of the instrument, along with its usage in similar type of studies dealing with people's perception/judgment of language use, was the main reason for selecting this instrument to determine whether testers' and not-testers' differ in their perception of the OPI as a speech event.

The SD instrument used in the study consists of 6 categories: *Format, Tester, Question, Candidate, Topic Nomination,* and *Turn-Taking*. For each major category, there are 4 scales. Thus, overall there are 24 scales on each SD form. Each scale includes a pair of opposing adjectives, such as "formal-informal," placed on a scale that ranges from 1 to 7.

In some instances, the sequence of the same adjectives has been reversed to prevent the repetition of the same pattern. Thus, if one major category included the pair of adjectives: "formal/informal," the order of the adjectives in the major category immediately following may have been reversed to read "informal/formal." The scales that include the same pairs of opposing adjectives reflect the same *aspects* (for example, the aspect of formality, naturalness, simplicity, cooperation, etc.). The following represents an example of the SD form used in the study

Semantic Differential:

An Example

The purpose of this study is to get some idea of your impression of the recorded samples of the oral test. In particular, we would like to know how you feel about the sample(s) you will hear. You will see that on each line there are two words expressing opposites.

For example, under ***FORMAT***

	A	B	C	D	E	F	G	
difficult	_____:	___:	_____:	_____:	_____:	_____:	_____:	**easy**

If you were to mark "A," it would mean that you think that the format of just heard sample was *extremely* difficult. If you were to mark "G," it would mean that you think that the format was *extremely* easy. The central position "D" indicates that you think that the format was *neutral*; that is neither easy nor difficult.

IMPORTANT:

(1) Please place your check-marks in the middle of the space, not on the boundaries
(2) Please check all the items
(3) Please never put more than one check-mark.

Please go rapidly through all the items. It is your immediate impression in which we are interested.

PLEASE MARK YOU ANSWERS ON THE ANSWER SHEET.

FORMAT

	A	B	C	D	E	F	G	
01. formal	___	___	___	___	___	___	___	informal
02. conversation-like	___	___	___	___	___	___	___	interview-like
03. natural	___	___	___	___	___	___	___	contrived
04. spontaneous	___	___	___	___	___	___	___	controlled

TESTER

	A	B	C	D	E	F	G	
05. spontaneous	___	___	___	___	___	___	___	controlled
06. uncooperative	___	___	___	___	___	___	___	cooperative
07. active/involved	___	___	___	___	___	___	___	passive/uninvolved
08. formal	___	___	___	___	___	___	___	informal

QUESTIONS

	A	B	C	D	E	F	G	
09. varied	___	___	___	___	___	___	___	repetitive
10. unnatural	___	___	___	___	___	___	___	natural
11. conversation-like	___	___	___	___	___	___	___	interview-like
12. formal	___	___	___	___	___	___	___	informal

CANDIDATE

	A	B	C	D	E	F	G	
13. tense	___	___	___	___	___	___	___	relaxed
14. formal	___	___	___	___	___	___	___	informal
15 active/involved	___	___	___	___	___	___	___	passive/uninvolved
16. uncooperative	___	___	___	___	___	___	___	cooperative

TOPIC NOMINATION

	A	B	C	D	E	F	G	
17. interview-like	___	___	___	___	___	___	___	conversation-like
18. negotiated	___	___	___	___	___	___	___	controlled
19. repetitive	___	___	___	___	___	___	___	varied
20. formal	___	___	___	___	___	___	___	informal

TURN-TAKING

	A	B	C	D	E	F	G	
21. varied	___	___	___	___	___	___	___	fixed
22. controlled	___	___	___	___	___	___	___	uncontrolled
23. formal	___	___	___	___	___	___	___	informal
24. conversation-like	___	___	___	___	___	___	___	interview-like

PROCEDURES

Collection of the SD data took approximately five weeks. The author met with the eight participants on an *individual* basis two or sometimes three times a week for an hour (with the exception of the first session that lasted an hour and a half). Each of these individual sessions proceeded in the following manner. First, the participants were asked to read the introduction to the SD. Second, they listened to the selected portion of the OPI. Third, they filled out the SD form. Each participant listened to 16 OPIs (i.e., 4 for each base level) and filled out 48 SD forms.

It is important to note that when asked to listen to a portion of an OPI, the participants were *not* informed that they were listening to the *Warm-up* or the *Level Check* phase of the OPI. However, some testers recognized the phases of the OPI, saying "O.K. now it is the *Wind-down*." The author, who prepared the appropriate SD forms, marked the backs of the SD forms in order to be able later on to distinguish between various phases of a given OPI.

DATA ANALYSIS

In the SD data analysis, an attempt was made to determine (i.e., to compare) whether there was a significant difference between testers' and non-testers' judgment of each individual scale across all 4 levels of speaking proficiency. That is, all individual scales within the six categories such as *Format, Tester, Topic* etc. were compared (i.e., 24 scales within the *Warm-up*, 24 scales the *Level Check* phase, and 24 within the *Wind-down* phase across all four levels of speaking proficiency were compared). A two-way analysis of variance was performed. This statistical procedure was selected due to the fact that in this factorial design there are two independent variables: (a) *Groups* (i.e. TESTER), with two levels, Testers and Non-Testers; and (b) *Level* of speaking proficiency with four levels: 1, 2, 3, 4. Where the ANOVA F ratio was significant, a Student Newman Keuls post hoc statistical procedure was performed.

RESULTS

The Warm-up Phase

Testers and non-testers differ significantly as to the degree of the following categories:

- Format

Testers perceive the format of the *Warm-up* as more conversation-like than non-testers:

TABLE 1 **FORMAT**

ANOVA—Individual Scales: *Warm-up* scale 2

Source	DF	AnovaSS	MeanSquare	FValue	Pr>
TESTER	1	12.50000000	12.50000000	3.93	.0498*
LEVEL	3	52.12500000	17.37500000	5.46	.0015*
TESTER*LEVEL	3	2.25000000	0.75000000	0.24	.8714

Student-Newman-Keuls test

SNKGrouping	Mean	N	TESTER
A	3.4688	64	N
B	2.8438	64	T

Student-Newman-Keuls test

SNKGrouping	Mean	N	LEVEL
A	4.1875	32	1
B	3.1875	32	2
B			
B	2.6250	32	3
B			
B	2.6250	32	4

- Questions

Testers judged questions posed within this phase as more natural than did non-testers:

TABLE 2 **QUESTIONS**

ANOVA—Individual Scales: *Warm-up* scale 10

Source	DF	AnovaSS	MeanSquare	FValue	Pr>
TESTER	1	8.50781250	8.50781250	4.11	.0448*
LEVEL	3	2.71093750	0.90364583	0.44	.7270
TESTER*LEVEL	3	0.58593750	0.19531250	0.09	.9630

Student-Newman-Keuls test

SNKGrouping	Mean	N	TESTER
A	5.7500	64	T
B	5.2344	64	N

- Topic Nomination

Non-testers perceived *Topic* nomination pattern to be more interview-like than did testers:

TABLE 3 **TOPIC NOMINTATION**

ANOVA—Individal scales: *Warm-up* scale 17

Source	DF	AnovaSS	MeanSquare	FValue	Pr>
TESTER	1	33.00781250	33.00781250	8.08	.0053*
LEVEL	3	14.39843750	4.79947917	1.17	.3224
TESTER*LEVEL	3	4.21093750	1.40364583	0.34	.7939

Student-Newman-Keuls test

SNKGrouping	Mean	N	TESTER
A	4.9844	64	T
B	3.9688	64	N

- Turn-Taking

Testers judged turn taking distribution within the *Warm-up* more varied (Table 4), and more conversation-like (Table 5) than non-testers:

TABLE 4 **TURN TAKING**

ANOVA—Individual Scales: *Warm-up* scale 21

Source	DF	AnovaSS	MeanSquare	FValue	Pr>
TESTER	1	48.75781250	48.75781250	12.39	.0006*
LEVEL	3	14.71093750	4.90364583	1.25	.2962
TESTER*LEVEL	3	3.27343750	1.09114583	0.28	.8417

Student-Newman-Keuls test

SNKGrouping	Mean	N	TESTER
A	4.2031	64	N
B	2.9688	64	T

TABLE 5 **TURN-TAKING**

ANOVA—Individual Scales: *Warm-up* scale 24

Source	DF	AnovaSS	MeanSquare	FValue	Pr>
TESTER	1	15.12500000	15.12500000	4.59	.0342*
LEVEL	3	12.81250000	4.27083333	1.30	.2789
TESTER*LEVEL	3	1.56250000	0.52083333	0.16	.9243

Student-Newman-Keuls test

SNKGrouping	Mean	N	TESTER
A	3.4375	64	N
B	2.7500	64	T

The Level Check Phase

The groups differ significantly in their judgment of the following categories:

- Question:

Non-testers perceived questions as more formal than did testers:

TABLE 6 **QUESTIONS**

ANOVA—Individual Scales: *Level Check* Scale 12

Source	DF	AnovaSS	MeanSquare	FValue	Pr>
TESTER	1	10.12500000	10.12500000	4.34	.0394*
LEVEL	3	6.06250000	2.02083333	0.87	.4610
TESTER*LEVEL	3	1.68750000	0.56250000	0.24	.8676

Student-Newman-Keuls test

SNKGrouping	Mean	N	TESTER
A	3.4063	64	T
B	2.8438	64	N

- Topic Nomination:

Non-testers perceive the topic nomination of the *Level Check* as more interview-like than did testers (Table 7). Also, they differ significantly in their perception of topic negotiation. Non-testers perceived topic negotiation as more controlled than did testers (Table 8):

TABLE 7 **TOPIC NOMINATION**

ANOVA—Individual Scales: *Level Check* Scale 17

Source	DF	AnovaSS	MeanSquare	FValue	Pr>
TESTER	1	17.25781250	17.25781250	5.69	.0186*
LEVEL	3	9.58593750	3.19531250	1.05	.3713
TESTER*LEVEL	3	3.64843750	1.21614583	0.40	.7523

Student-Newman-Keuls test

SNKGrouping	Mean	N	TESTER
A	3.2188	64	T
B	2.4844	64	N

TABLE 8 **TOPIC NOMINTATION**

ANOVA—Individual Scales: *Level Check* scale 18

Source	DF	AnovaSS	MeanSquare	FValue	Pr>
TESTER	1	11.28125000	11.28125000	5.21	.0242*
LEVEL	3	11.93750000	3.97916667	1.84	.1440
TESTER*LEVEL	3	0.78125000	0.26041667	0.12	.9480

Student-Newman-Keuls test

SNKGrouping	Mean	N	TESTER
A	5.8281	64	N
B	5.2344	64	T

The Wind-down Phase

The results of the SD data analysis indicate that there are significant differences between testers' and non-testers' perception as to the degree of the following individual scales of the *Wind-down* phase:

- Format

Non-testers perceived the format of the *Wind-down* phase as more informal than testers:

TABLE 9 **FORMAT**

ANOVA—Individual Scales: *Wind-down* scale 1

Source	DF	AnovaSS	MeanSquare	FValue	Pr>
TESTER	1	8.50781250	8.50781250	4.44	.0373*
LEVEL	3	12.89843750	4.29947917	2.24	.0870
TESTER*LEVEL	3	4.14843750	1.38281250	0.72	.5414

Student-Newman-Keuls test

SNKGrouping	Mean	N	TESTER
A	5.8906	64	N
B	5.3750	64	T

Non-testers judged questions as being less formal than testers:

TABLE 10 **QUESTIONS**

ANOVA—Individual Scales: *Wind-down:* scale 12

Source	DF	AnovaSS	MeanSquare	Fvalue	Pr>
TESTER	1	9.57031250	9.57031250	4.07	.0458*
LEVEL	3	12.02343750	4.00781250	1.71	.1695
TESTER*LEVEL	3	1.83593750	0.61197917	0.26	.8537

Student-Newman-Keuls test

SNKGrouping	Mean	N	TESTER
A	5.7031	64	N
B	5.1563	64	T

SUMMARY OF THE SEMANTIC DIFFERENTIAL DATA ANAYSIS FINDINGS

The findings the SD data analysis show that in the perception of both groups (i.e., Testers and Non-testers), the *Level Check* differs substantially from the other two phases: the *Warm-up phase* and the *Wind-down* phase. The **format** of the *Level Check* phase is viewed by both testers and non-testers as **formal**, **interview-like**, and **controlled**. **Topic nomination** within the *Level Check* is perceived by both groups as: **interview-like**, **very controlled**, and **formal**. The **turn-taking mechanism** in the *Level Check* is perceived by both testers and non-testers as: **controlled**, **formal**, and **interview-like**. The participants in the two groups agree that the tester's verbal behavior within this phase is even less spontaneous and more formal that in the other phases.

Testers' and non-testers' perception of the *Warm-up* and the *Wind-down* differs from their perception of the *Level Check*. In contrast to the *Level Check*, the *Warm-up* phase is perceived by **both** testers and non-testers as: **more conversation-like, more natural, more spontaneous**, and **varied** in terms of the format, topic nomination, and question types. However, testers judged this phase as being more conversation-like than non-testers. This should not be surprising since OPI testers are trained to consider this phase very conversation-like in nature. What seems to be significant, however, is that both testers and non-testers perceive turn-taking distribution to be controlled, which points toward the kind of asymmetry of power typical of an interview in general. Also, *Level 1* is significantly different from other levels in the

perception of both testers and non-testers. In comparison to other levels, the format of the *Level 1 Warm-up* is perceived by all SD participants as the least conversation-like. The candidate's level of speaking proficiency seems to have some impact on how the SD participants perceive the format of the *Warm-up* phase.

Testers' and non-testers' perception of the *Wind-down* phase is similar to their perception of the *Warm-up* phase. The findings of the SD analysis show that similar to the *Warm-up* phase, both groups perceived this phase as being **more conversation-like**, **more informal** in terms of the format, question types, and topic nomination. However, testers perceived the format of the *Wind-down* as less natural and less spontaneous than non-testers. Also, testers perceived the questions posed within this phase as less informal than non-testers. Again, the OPI testers training may offer some explanation to this difference in perception between testers and non-testers. OPI testers are trained to view the *Wind-down* phase as "the interviewer's last chance to check any aspect of the candidate's ability that may still be incompletely assessed" (ETS, 1992, p. 24). For the "outsiders"—non-testers, the format and type of questioning may give an impression of being casual and spontaneous, but to testers, the *Wind-down* phase allows assessment of the candidate's speaking ability.

DISCUSSION

THE LEVEL CHECK PHASE

In the native speakers' judgment, the *Level Check phase* of the OPI represents a very formal type of interview, in which not only does the tester control when and for how long the candidate will hold the floor, but also the topic the candidate will talk about. The role of the tester/interviewer is to provide a stimulus (a question or a task), and the role of the candidate/interviewee is to respond. There is little negotiation allowed on the part of the candidate. The following example may offer some explanation for the native speakers' perception of the *Level Check* as a very formal type of interview. (Please note that the numbers do not correspond to lines but *turns*, and that the first number indicates the level. Thus, for example, 209 means level 2 tape 9):

244. INTER: O.K. let's imagine this say he asks you you talked you tell me you talked about immigration I don't understand that word. Explain that to me daddy.

245. CAND: Well I don't think my son is uh at the stage where he knows about immigration

yet.

246. INTER: Is there any way you can think of for me that that you might try to tell him what immigration is?

247. CAND: At three years old? z

248. INTER If he hears you say that word how would you tell him or you could say you could exp.. z

249. CAND: Yeah I know what you mean like if he keeps uttering that word all

the time z

250. INTER: Yeah z

251. CAND: and he wants to know the meaning of [c] immigration?

252. INTER: Yeah!

253. CAND: Although I'm not assuming that he will utter that word all the time because it well let's assume that he is uttering so I would always explain to him immigration as as his father his father immigrated to this country. I came from a different country to uh this country or immigration *is* like to come from another country into another country.

This excerpt illustrates the candidate's limited power of negotiation of a new topic/task. Despite his objection as to the "validity" of the task presented to him (i.e., the candidate clearly finds its improbable that his three year old son will ask for the explanation of the word immigration), the candidate is forced to respond. His comments regarding the credibility of the task are being totally ignored by the tester who acts as if he has a right to "do" whatever he finds important to his own agenda. He will not change the task even though the candidate's objections make sense because it would mean giving up his power.

Also, question posed by testers in the *Level Check* phase may provide some explanations for the native speakers' decision to view the *Level Check* as a very formal behaviorist type of interview. As an illustration consider the *Level Check* tape 402. After the discussion of the topic of buying a new car (Turn 117), the tester introduces the topic of immigration (Turn 137), which is followed by the topic of militant groups (Turn 139), which is followed by the topic of education (Turn 145). These questions are so formal and so unrelated that they sound as if they had been prepared ahead of time; as if they were being read by

the tester:

117. **INTER**: Uh-huh. OK! Good. Uh **if you: were to buy a car .. uh what kind of a car would you get?** Would you buy an expensive one a beautiful one or a well-made one?

137. **INTER:** Uh-huh. Good. OK. Very good. Uh next question. Uh we hear a lot about immigration these days in the news .. uh many problems many uh uh comments by different people. Uh immigrants themselves feel that in this country they're not treated well. And they feel like they are victims. Uh could you tell me why you think uh: uh immigrants feel like they are victims in this country? They were victims where they came from and [c] now when they come here they feel like victims. Could you tell me why you think they feel that way?

139. **INTER**: Uh-huh. OK. Uh did you read this uh weekend about that those militant groups uh in our country who uh claim to dislike uh: our government and they feel that uh uh: they uh now these are not immigrants but they're unhappy with the situation [c] with the government here. Uh I I was thinking uh do you think that people should stay in this country if they're unhappy with the government? **I'm talking about these militant people [Laughs]. Do you think they should stay if they stay?**

145. **INTER**: It's interesting to hear their comments. Uh OK: uh uh **twenty-five years ago there was a court case called Brown versus the Board of Education and the phrase that they used a lot was "separate but equal" Uh do you know what that means?**

Some testers often explicitly tell the candidate that they do not have to "tell the truth" while expressing their opinions about a given topic. This stands in a sharp contradiction to what participants in a real conversation are expected to do, and, therefore, may provide some additional support for the native speakers' decision to judge the *Level Check* as a very formal interview and not as everyday conversation. The following excerpt from tape 205 illustrates this point:

170. **INTER:** Uh next question. O.K. uh this is on abortion O.K. **you don't have to tell me your real views on this O.K. just make them up.** In your opinion what circumstances should exist before a physician or a doctor suggests to his patient that she should consider an abortion.

In the *Level Check* the tester makes frequently evaluative remarks prior to introducing a new topic/task, or a role play situation. These explicit remarks serve as a reminder to the candidate that this is

not a conversation. The following excerpt from tape 303 illustrates this point:

> **69. INTER: Let me ask you this question** also if I may and I want to emphasize for you before you even answer that it is asking going a little bit into your personal life but I am not prying into that. I have a **linguistic reason** why I want to ask you this question and here it is. We are of course at the beginning of the week now this is Monday and we've just come off the weekend. Could you briefly tell me how you spent the weekend starting from Friday night or Saturday morning until last night.

The phrase "Let me ask you this question," combined with "I have a linguistic reason" sends a signal to the candidate that he is *not* being engaged in a causal conversation, but that his language skills are being assessed.

The *Level Check* phase is the most important for the process of the OPI rating. Some testers even make very explicit remarks to order to separate this phase from the *Warm-up* phase. The following portion of the OPI, tape 303, illustrates this point. After a very lengthy *Warm-up*, which centers on the topic of the weather in Mexico City, the tester gives an explanation of the agency testing procedures where he explicitly states that what has been discussed up to this point (i.e., up to the *Level Check* phase) is apparently irrelevant. That is, *Warm-up* phase seems to have little impact on the outcome of the OPI final rating:

> **59. INTER: You and I sir are going to be talking in English for just a while a little while this afternoon** and after we finish I'll turn you over to uh what we'll do is we'll hang up and then then you'll be speaking in your foreign language conversation for a while. But concerning both tests let me give you a little bit of information a little bit of information if I may.

The tester's remark: "You and I sir are going to be talking in English for just a while a little while this afternoon" sounds out of place. It contradicts what they have been doing so far. They have been talking in English for some time now, but apparently it does not have any effect on the candidate's final rating. The "real" talk in English will begin as of the next phase—*Level Check*. The tester's remarks seem to serve as a signal to the candidate. As of now, the candidate needs to be careful how he speaks in English.

The native speakers' judgement about a very control nature of turn-taking and topic nomination

points to the asymmetrical distribution of power typical of an interview. This asymmetry is evident in the tester's exclusive control over when and for how long a speaker holds the floor, what topics are to be discussed, and what is, or is not relevant to the interview. The tester has the power to interrupt the candidate in any place he/she considers relevant as turn 135 in tape 409 illustrates:

> **135. CAND:** I would certainly think that the first thing that I would do is set some sort of controls to avoid any any guns or weapons in school that would be my very first step to take. Second I would immediately implement as part of the school uh: schedule one course uh uh maybe daily half an hour a day or one hour a day or or whatever to the school's schedule uh: on educating children about weapons and about violence and giving them statistics and showing them uh movies and the same way that these children are being taught to be violent how carrying weapons to school how gangs are cool well we should show them also the other side of the coin I believe this should be also part *of* the school's program. I think it should be part of the education and I think the only way children are going to: stop imitating is if they understand we can't force anything into anyone and nobody learns with somebody else's z
>
> **136. INTER:** **O.K. let's just shorten it to right here** and switch gears just a little bit same subject only now you're talking to your son.

In turn 135, the candidate responds to the task presented to her by the tester in the preceding turn where she was asked to address the school board regarding the subject of safety in schools. Her argument for the ban of guns or weapons in schools is interrupted by the tester in turn 136: "O.K. lets just shorten it right here and switch gears just a little bit," and she is given another task.

THE WARM-UP and WIND-DOWN PHASES

The findings of the study indicate that the *Warm-up* and the *Wind-down phases* differ in the native speakers' judgment from the *Level Check* phase. However, although according to the native speakers these phases contain more conversation-like features, the controlled nature of turn-taking, and lack of topic negotiation prevent these two phases from being viewed as everyday conversation. The combination of conversation-like features and interview-like features exhibited in the *Warm-up* and *Wind-down* seem to point in the direction of a more conversation-like type of an interview; a type similar to a *sociolinguistic interview* (Mishler, 1986).

speaking ability in the context of an interview. To be more precise, it tests speaking ability in the context of two types of interviews: a very formal type of interview that exhibits many features of a survey research interview, which is based on the behaviorist theory of stimulus and responses, and a more conversational type of interview that exhibits many features of a sociolinguistic interview (Mishler, 1986).

The native speakers' perceptions of the OPI communicative speech event contradict the ETS's claim that "a well-structured oral proficiency interview tests speaking ability in a real-life context—a conversation. It is almost by definition a valid measure of speaking ability" (ETS, 1989). Such a contradiction undermines the validity of the OPI testing instrument that impacts the users' ability to generalize scores from a testing context to the outside world. For example, since the OPI claims to be conversational in nature, the users of the OPI may be under the impression that the candidate who obtained a level 2 or higher is able to fully participate in a conversation. That is, the candidate is able to "compete" for the floor, negotiate a new topic, etc. The findings of the study show that the candidate does not have many opportunities to prove the mastery of these skills. The phases in which the candidate may exhibit more initiative—the *Warm-up* and the *Wind-down*—are minimized in the process of assigning a final rating. The final rating is based on the candidate's performance within the *Level Check*—the phases wherein the candidate's power to negotiate is almost nonexistent.

From a practical point of view, the findings of the study may be used to improve testers' elicitation techniques, especially within the *Level Check* phase of the OPI. In the OPI training workshops, as an exercise, testers could be asked to listen to samples of taped natural conversations, and to discuss major characteristics of everyday conversation. Improvement is also urgently needed regarding the so-called lead-in questions (i.e., questions that lead to a given task). Improvement in this area might alleviate the impression that the testers are working from a prescribed set of questions based on their current reading of political and economic news. Testers should also be encouraged to avoid making any evaluative remarks such as: "Very good, let's move to the next question." " We made it. Good, very good." "Uh-huh. Good. OK. Very good. Uh next question." "OK. I have asked you all the questions from my list." Such remarks clearly point toward a test and *not* a casual conversation.

This study recommends further research in the area of the construct validity of the OPI and other oral proficiency performance tests. The investigation of construct validity should be based, as Messick

In sociolinguistic interviews, the interviewee has greater control over who holds the floor. The interviewee is allowed to ask questions and negotiate a new topic. This contributes to a better sense of cooperation and solidarity between the interviewer and the interviewee, and thus makes the asymmetry of power less evident. Overall, in sociolinguistic interviews the interaction is more locally managed. That is, it is managed on a turn-by-turn basis with the interviewer more closely adhering to what has been said previously. The following example illustrates these points in support of the native speakers' perception of the *Warm-up* as more conversational type of interviews:

25. INTER: Oh I see, I *see*, so that's your native *language*?

26. CAND: Yes.

27. INTER: Oh uh .how long have you *been* in the United States?

28. CAND: Since 1981.

29. INTER: Are you in the Virginia suburbs r the Maryland suburbs?

30. CAND: I'm in Virginia.

31. INTER: Oh in Virginia, so am I.. I live in Fairfax County.

32. CAND: Me too. I live in uh west of Springfield.

33. INTER: Oh that is more or less the area I live in. I'm in Burke.

34. CAND: We are right close to Burke, anyway.

In this excerpt taken from tape 209, the tester offers some personal revelations. Both the tester and the candidate contribute to the topic that is locally managed. Also, the pattern of *question/response* is not so prevalent as in the *Level Check* phase.

CONCLUSION

The findings of the *Semantic Differential* provide some answers to the research question:

> *Do native speakers (testers versus non-testers) differ in their judgments of the nature of the OPI communicative speech event?*

With a few exceptions where testers and non-testers differ as to **the degree** of certain aspects within the *Warm*-up, *Wind-down*, and *Level Check* phases of the OPI, the native speakers do **NOT** differ in their judgments of the nature of the OPI communicative speech event. In the native speakers' judgment, the OPI does not test speaking ability in the real-life context of a conversation. In their perception, the OPI tests

(1989) suggests, on empirical evidence and theoretical rationales. The empirical evidence should provide detailed information about **what** participants actually do in the OPI (i.e., the information about the discourse structure of the OPI), and **how** people perceive/judge what it is that has been done in the OPI (as this study has attempted to investigate). The theoretical rationale, on the other hand, should be based on the recent findings in the field of interaction analysis, conversation analysis, and ethnography of communication. The time has come for the OPI to concentrate some efforts on improving its construct validity. Research into the nature of the OPI's construct validity may not only improve the quality of the OPI testing instrument, but may prove to be indispensable for designing better instruments for assessing language speaking ability.

REFERENCES

Bachman, L. F. 1990: *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F. 1988: Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition 10*, 149–174.

Bachman, LF. and Savignon, S. 1986: The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal 70* (4), 380–390.

Brown, G. and Yule, G. 1983: Discourse *analysis*. Cambridge: Cambridge University Press.

Clark, J. L. D. 1972: *Foreign language testing: Theory and practice*. Philadelphia, PA: Center for Curriculum Development.

Clark, J. L. D. and Clifford, R. 1988: The FSI/ILR/ACTFL proficiency scales and testing techniques: development, current status, and needed research. *Studies in Second Language Acquisition 10*, 129–147.

Educational Testing Service. 1989: *Oral proficiency testing manual*. Princeton: Educational Testing Service.

Educational Testing Service. 1992: TOEIC Language Proficiency Interview Manual. Princeton, NJ: Educational Testing Service.

Hughes, A. 1989: *Testing for language teachers*. Cambridge: Cambridge University Press.

Johnson, M. 1997: *What kind of speech event is the Oral Proficiency Interview: Problems of construct validity*. Unpublished Georgetown University Disseration.

Johnson, M. 1997: *Interaction in the Oral Proficiency Interview*. Paper presented at the American Association for Applied Linguistics Annual Conference, Orlando, FL.

Johnson, M. and Tyler, A. 1998: Re-analyzing the OPI: How much does it look like natural conversation? R. Young and He A.W., editors, *Taking and Testing: Discourse Approaches to the Assessment of Oral Proficiency*. Amsterdam: John Benjamins.

Jones, E. E. and Gerard, H. B. 1967: *Foundation of social psychology*. New York: Wiley and Sons.

Lantolf, J. and Frawley, W. 1985: Oral-proficiency testing: A critical analysis. *Modern Language Journal* 69 (40), 337–345.

Lantolf, J. and Frawley, W. 1988: Proficiency: Understanding the construct. *Studies in Second Language Acquisition* 10, 181–195.

Lazaraton, A. 1992: The structural organization of a language interview: A conversation analytic perspective. *System* 20, 373–386.

Levinson, S. 1983: *Pragmatics*. Cambridge: Cambridge University Press.

Liskin-Gasparro, J. 1984: The ACTFL proficiency guidelines: Gateway to testing and curriculum. *Foreign Language Annals* 17, 475–89.

Lowe, P. 1988: *ILR handbook on oral interview testing*. Washington, DC: DLI/LS Oral Interview.

Lowe, P. 1983: The ILR interview: Origins, applications, pitfalls, and implications. *Die Unterrichts Praxis 2*, 230-240.

Lowe, P. 1986: Proficiency: Panacea, framework, process? A reply to Kramsh, Schultz, and in particular, to Bachman and Savignon. *Modern Language Journal* 70 (4), 391-396.

McNamara, T. 1996: *Measuring second language performance*. London: Addisson Wesley Longman.

McNamara, T. 1996: *Interaction in second language performance assessment*. Plenary address. American Association for Applied Linguistics Annual Conference. Chicago IL, 26 March 1996.

Messick, S. 1989: Validity. Linn, R. L., editor, *Educational measurement* (pp. 13–104). New York: Macmillan.

Mishler, E. 1986: *Research interviewing. Context and narrative*. Cambridge: Harvard University Press.

Ross, S. 1992: Accommodative questions in oral proficiency interview. *Language Testing* 9, 173–186.

Ross, S. and Berwick, R. 1992: The discourse of accommodation in oral proficiency examinations. *Studies in Second Language Acquisition* 14, 159–176.

Sacks, H., Schegloff, E. and Jefferson, G. 1974: A simplest systematics for the organization of turn taking in conversation. *Language 50* (4), 696–735

Savignon, S. 1985: Evaluation of communicative competence: The ACTFL provisional proficiency guidelines. *Modern Language Journal* 69, 129–133.

Schegloff, E. and Sacks, H. 1973: Opening up closings. *Semiotica* 8, 287–327.

Schegloff, E., Jefferson, G. and Sacks, H. 1977: The preference for self-correction in the organization of repair in conversation. *Language* 53, 361–82.

Schiffrin, D. 1990: Conversational analysis. *Annual Review of Applied Linguistics* 11, 3–16.

Schiffrin, D. 1987: *Discourse markers*. Cambridge: Cambridge University Press.

Schiffrin, D. 1994: *Approaches to discourse*. Oxford: Blackwell Publishers.

Shohamy, E. 1982: Affective considerations in language testing. *Modern Language Journal 66*, 13–17.

Shohamy, E. 1983: The stability of oral proficiency assessment on the oral interview testing procedures. *Language Learning* 33, 527–40.

Shohamy, E. 1988: A proposed framework for testing the oral language of second/foreign language learners. *Studies in Second Language Acquisition* 10, 165–179.

Tyler, A. and Johnson, M. 1998: *Natural Context?: Collaboration, Conversational Involvement and the OPI*. Paper presented at the American Association for Applied Linguistics, Seattle, WA.

Valdman, A. 1988: Introduction to the assessment of foreign language oral proficiency. *Studies in Second Language Acquisition 10* (2), 121–128.

Van Lier, L. 1989: Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly* 23, 489–508.

Young, R. 1995: Conversational styles in language proficiency interviews. *Language Learning 45* (1), 3–42.

Marysia Johnson
Assistant Professor
Department of English
Arizona State University
P.O. Box 870302
Tempe AZ 85287-0302

Tel: 602-965-3810
Fax: 602-965-3451
Email: MarysiaJ@asu.edu